

# Help is on the WAIS

BY MARY LUKANUSKI

*The Wide Areas Information Servers (pronounced "ways") protocol will make navigating disparate databases easier, but not right away.*

**"I**s it control F to scroll forward? Control R? What do I use to print, Tab or p or P? Now what's the difference between OR and AND?" The anguish of having to remember multiple commands, remembering where the cheat sheet is located, and the idiosyncrasies of Boolean operators! That's the sad plight of the database searcher and the hopeless dilemma of the occasional user, going from one set of search commands to another as he or she changes databases. Database searching is one of the mysteries of our profession. Nonlibrarians may attempt to "search," but whom do they call on when they get frustrated or confused?

**Search a variety of databases through one interface? Before you say "Not!"—wait.**

Imagine the capacity to search a variety of databases through one interface. Imagine searching in everyday language, without having to use Boolean operators. Imagine not logging in and out when changing databases. Imagine accessing text, sound, and images with the same interface. Before you say "Not!"—wait. This is all being done within the Wide Areas Information Servers (WAIS) protocol.

MARY LUKANUSKI is data collection librarian at Rand, the not-for-profit research and policy institute, in Santa Monica, Calif.

The WAIS (pronounced "ways") protocol began as a project initiated by Thinking Machines, a firm that designs massive parallel computers and software for their machines. Thinking Machines also involved Apple Computer, Dow Jones & Company, and KPMG Peat Marwick. The goal of the project was to create a system that would allow a user to access and manipulate personal, corporate, and commercial information through one interface. Thinking Machines provided the software and hardware. Apple concentrated on the interface. Dow Jones News Retrieval permitted use of its database, and Peat Marwick served as a test site.

intricacies of the reference interview, Kahle came across the ANSI-NISO Z39.50 standard, the common language used by online databases. Kahle and other project members decided the Z39.50 standard could serve as a model for the common language between WAIS clients and WAIS servers.

After a year of development, a system evolved that allowed users access to personal, corporate, and published information—such as an online database—from one interface. Kahle describes it as a personal publishing tool. The icon-driven interface will be familiar to those who use Macintosh computers or the

**Nonlibrarians may attempt to "search," but whom do they call on when they get frustrated or confused?**

Before work began on the project, it was decided that WAIS would be a client-server protocol in which the client is the requester and the server is the provider of the database.

## Savage user interfaces

When designers from Apple's Advanced Technology Group began thinking about the interface between software and user, they consulted two librarians from the Group library, Janet Watts and Steve Cisler, to see what other database interfaces existed. Cisler and Watts demonstrated several online databases. "We showed them how bad it was," said Watts. "Savage user interfaces," echoed Cisler.

Working with Cisler and Watts, the designers realized they were attempting to reproduce the human interface in information gathering. They became intrigued by the concept of the reference interview and how the reference librarian determines what the user wants.

Brewster Kahle, cofounder of Thinking Machines and leader of the WAIS project, was also intrigued—enough so that he enrolled in reference classes at the Simmons College Graduate School of Library and Information Science. While studying the

Windows environment and employs a "question box," a "source box," and an "answer box." The user poses a natural language query such as "What was the rate of inflation last year in the United States?" in the question box. Notice the absence of Boolean operators. Next, the user selects the sources listed in the source box: news services, financial services, in-house technical reports—whatever is available online.

The user selects a source and the query is posed to the source by matching words or phrases that appear in both the query and the source. Matches pop up in the answer box. The user can then read a brief description of the items retrieved, select items for viewing in a fuller form, or ask for other relevant documents.

## On the Internet

Once WAIS was completed, Kahle continued to pursue the wide area concept in a wider arena—the Internet. While Apple, Dow Jones, and Peat Marwick continue to be involved in other capacities, Kahle and Thinking Machines are promoting the potential of WAIS. The company is offering the protocol software free via the Internet, and Kahle is heavily engaged in WAIS dis-

cussions with the networked information community.

Reactions to the WAIS protocol are varied. Nonlibrarians are enthusiastic. Database searching is no longer intimidating, and personalized information can easily be found without the intermediary of a reference librarian. The intermediaries, however, are less than enthusiastic. "Professional searchers have been suspicious," commented Apple's Watts on librarians' reactions to the WAIS search capacities. "They have less control over the search and feel a need to understand how it works." The broader library community is just beginning to discover WAIS, and judging from the conferences and workshops on the subject, WAIS is engendering a great deal of interest.

---

### **Over 100 databases and 5,000 individuals are now using the WAIS protocol.**

---

A primary benefit of the WAIS protocol is that it is one method of "navigating the network." An overwhelming amount of information is currently available to anyone with a PC and a modem. Through the use of natural language querying and relevant document recall, WAIS offers a very easily understood method of accessing resources on the Internet. Electronic newspapers, tailored to individual taste, would be accessible. Resources such as picture libraries, OPACs, corporate libraries, and electronic text libraries would be available to anyone with access to the Internet.

The result is that many purveyors and seekers of information have been attracted to WAIS. According to Kahle, over 100 databases and 5,000 individuals are now using the WAIS protocol. The Library of Congress is planning to make its catalog available through WAIS. Dow Jones, involved in the project since its inception, will use the WAIS protocol on its Dow Vision network, which will contain the *Wall Street Journal* and 450 other business-related publications.

This grand vision is not without problems or criticisms. Affecting all resources linked by the Internet is the uncertainty of federal funding. Although the National Research and Education Network (NREN), which would expand and improve Internet, has been approved, the project has yet to be funded.

Currently, the protocol requires a power-

ful search engine. WAIS, which uses the UNIX operating system, runs on two massive parallel computers, the Connection Machine 2 and Connection Machine 5, which are produced only by Thinking Machines. These units are performing well; however, how they will respond to increased demand is uncertain, as is the future of very large parallel computers in an age of distributed computing.

#### **A WAIS to go**

Security, attracting additional commercial vendors, and the pricing of information are also matters of concern. Solutions to the problem of security in a network environment—in the form of varieties of encrypting packages—abound, but no consensus has been made on which, if any, of these packages should be used. Additionally, it should be noted that the federal government is openly nervous about the existence of encrypting packages. Senate bills S266 and S618, which concern terrorism and violent crime, both state: "It is the sense of Congress that providers of electronic communications services... shall ensure that communications systems permit the government to obtain plain text contents of voice, data, and other communications when appropriately authorized by law."

The economic health of WAIS is also a matter of concern if the protocol is to become viable. If WAIS is to expand and reach its potential, commercial vendors will have to be attracted to using the protocol. Internet is for research and academic use, not for commercial vendors. The some 100 databases now available through WAIS are fun and interesting, but they don't pack the same economic punch as DIALOG would if it adopted the protocol. A related concern is the pricing of information. How users will be charged and what those charges will be is speculated, but as of yet unstated. Dow Vision will be the first for-fee server using the protocol. How Dow Jones handles Dow Vision and user response to Dow Vision undoubtedly will influence other commercial vendors.

---

### **Dow Jones... will use the WAIS protocol on its Dow Vision network.**

---

George Brett, program manager for the Networked Information Center for Communication of the Microelectronic Center

---

### **Designers realized they were attempting to reproduce the human interface in information gathering.**

---

for North Carolina, believes that the notion of networked information was "So what?" and the question has now changed to "Now what?" We all know the grand vision of networked information—how rich resources will be available to anyone with a PC, modem, and a credit card. How we make the most of these resources is the challenge we face now. So, hold on to those cheat sheets; information at your fingertips may have a WAIS to go. □

For further information on WAIS, contact Brewster Kahle through Internet: Brewster@Think.com

---

# SCIENCE NEWS

June 22, 1991  
Vol. 139, No. 25  
Pages 385-400

## Computers

*Ivars Peterson reports from San Jose, Calif., at the Physics Computing '91 conference*

### Navigating the information swamp

The ubiquitous lab notebook, with its dog-eared corners, stained pages and scribbled entries, may one day give way to an electronic analog that permits not only the recording of data but also the sharing of information among researchers scattered throughout the world. Researchers at Baylor College of Medicine in Houston have developed a sophisticated, computer-based scheme, called the Virtual Notebook System, that allows its user to gather, organize and annotate information selected from a variety of sources.

With such a notebook, a medical researcher interested in the diagnosis of a certain ailment, for example, can readily assemble a package consisting of X-ray images, personal comments, citations, journal articles, news items, electronic-mail extracts and other relevant pieces of information. Moreover, the researcher can instantly share that information with others who use the same system, even if they are thousands of miles away. "You can even write in someone else's notebook," says Kevin B. Long, who directed the project.

Designed to facilitate collaboration, the system's key element consists of software that masks the underlying maze of computers and computer networks that often stands in the way of efficient and convenient communication among researchers working with different computer equipment. The Virtual Notebook System also incorporates a new programming approach for simplifying the indexing and retrieval of information stored in computers. A specially programmed, information-seeking computer — known as the Wide Area Information Server and developed under the direction of Brewster Kahle of Thinking Machines Corp. in Cambridge, Mass. — responds to requests typed in English. Users don't have to know exactly how to find the information they need; nor do they have to remember any special instructions to locate data.

Best suited for groups of researchers already linked by computer networks, the Virtual Notebook System may prove a crucial component of large collaborative efforts. Officials with the Superconducting Super Collider are investigating the system as a possible means of sharing and analyzing experimental data when the accelerator is eventually completed.